

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-123406

(43)Date of publication of application : 26.04.2002

(51)Int.Cl.

G06F 11/20  
G06F 3/06  
G06F 12/16

(21)Application number : 2000-316213

(71)Applicant : PFU LTD

(22)Date of filing : 17.10.2000

(72)Inventor : KAWAI KEISUKE

SHIMADA KAZUHIRO

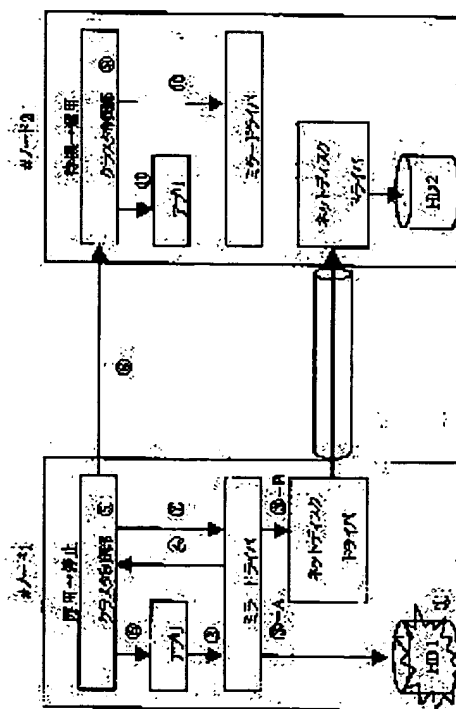
IMAI HIDEKAZU

## (54) HIGH RELIABILITY SYSTEM

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To suitably deal with fault occurrence when employing a mirror ring configuration with the disk of an active system node as an original and the disk of a reserve system node, which is connected through a network, as a shadow.

**SOLUTION:** When a fault occurs on the disk of the active system node, a node operated as a reverse system is turned into active state and the node operated up to the moment as an active system is stopped so that processing can be executed to stop operation as a duplex system. Besides, when a fault occurs on the disk of the reserve system node, the node operated as a reserve system is stopped so that processing can be executed to stop operation as a duplex system. Besides, when a fault occurs on a network prepared for access to a shadow disk, the node operated as a reserve system is stopped so that processing can be executed to stop operation as a duplex system.



## LEGAL STATUS

[Date of request for examination] 14.10.2003

[Date of sending the examiner's decision of rejection] 29.11.2005

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

**\* NOTICES \***

**JPO and NCIPi are not responsible for any damages caused by the use of this translation.**

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

**CLAIMS**

---

[Claim(s)]

[Claim 1] It has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network. present -- business -- In the high reliability system which carries out business by using the disk which is normal, and which is already one side when a failure occurs on one of disks When a means to detect whether it is that the failure occurred on the above-mentioned original disk, and failure generating of the above-mentioned original disk are detected it corresponds to the means which changes a standby system node to an employment condition, and the above-mentioned node change -- making -- till then -- present -- business -- the high reliability system characterized by having a means to stop duplex system by stopping the node which was operating as a system node.

[Claim 2] It has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network. present -- business -- In the high reliability system which carries out business by using the disk which is normal, and which is already one side when a failure occurs on one of disks The high reliability system characterized by having a means to detect whether it is that the failure occurred on the above-mentioned shadow disk, and a means to stop duplex system by stopping a standby node when failure generating of the above-mentioned shadow disk is detected.

[Claim 3] It has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network. present -- business -- In the high reliability system which carries out business by using the disk which is normal, and which is already one side when a failure occurs on one of disks present -- business, when a means to detect whether it is that the failure occurred to the network for shadows prepared in order that a system node may access the above-mentioned shadow disk, and failure generating of the above-mentioned network for shadows are detected The high reliability system characterized by having a means to stop duplex system by stopping a standby node.

---

[Translation done.]

\* NOTICES \*

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] this invention -- present -- business -- it is normal, when it has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network and a failure occurs on one of disks -- already -- while is related with the high reliability system which enables it to cope with it suitable for failure generating especially about the high reliability system which carries out business by using a disk.

[0002]

[Description of the Prior Art] The high reliability system which consists of a server which performs business, and a server which stands by in preparation for the failure of the server is used widely. the former -- present -- business -- it is called a system node and the latter is called the standby system node.

[0003] present -- business -- when a system node is downed according to a certain failure, a standby system node takes over business. in order to take over this business -- operating data -- present -- business -- it is necessary to take over to a standby system node from a system node

[0004] a this taking over of former and operating data sake -- present -- business -- the configuration of preparing the common disk unit connected with a SCSI interface etc. between a system node and a standby system node was taken.

[0005] However, a common disk unit is very expensive. then -- present -- business -- it is normal, when it has the mirror set function (function of RAID1) which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network and a failure occurs on one of disks -- already -- the high reliability system which takes the configuration of carrying out business by using a disk is being used for while.

[0006] this high reliability system shows to drawing 5 -- as -- present -- business -- when writing operating data in an original disk, a system node is writing that operating data in a shadow disk using a network disk driver, and has taken the configuration of performing mirroring.

[0007] and -- present -- business -- the system node has taken the configuration of carrying out business by accessing a shadow disk using a network disk driver, when business is carried out by accessing an original disk when an original disk is normal and a failure occurs on an original disk.

[0008]

[Problem(s) to be Solved by the Invention] However, there are the following troubles in this conventional technique.

[0009] namely, the time of a failure occurring on an original disk -- present -- business -- since a system node performs I/O of operating data to a shadow disk altogether, the entry-of-data engine performance falls and it has the trouble that system-wide degradation occurs [ this ] to accumulate.

[0010] Moreover, when a failure occurs on a shadow disk, the configuration of stopping a standby system node is not taken with this conventional technique.

[0011] after a failure occurs on a shadow disk from now on -- present -- business -- the time of the system node carrying out business using the original disk -- present -- business -- if it happens that a system node carries out a system down, as for a user, business should be taken over to a standby system node -- it is that come out, and it is not performed although it is, and there is a trouble of having anxiety to the dependability of a system.

[0012] and -- this -- present -- business -- since the operating data of that shadow disk are not updated even if it may become possible to use a shadow disk by avoiding a failure part in the case of the system down of a system node, a \*\*\*\*\* system node (node with that shadow disk) also has the trouble that the newest operating data cannot be succeeded.

[0013] moreover -- present -- business -- also when a failure occurs to the network (exclusive network of drawing 5 ) prepared in order that a system node may access a shadow disk, the configuration of stopping a standby system node is not taken with this conventional technique.

[0014] From now on, a trouble in case a failure occurs, and the same trouble will be in a shadow disk.

[0015] The local disk of a system node is used as an original disk. this invention is made in view of this situation -- having -- present -- business -- It has the mirror set function which uses as a shadow disk the local disk of a standby system node connected through a network. When a failure occurs on one of disks, it is, when taking the configuration of carrying out business by using the disk which is normal and which is already one side, and aims at offer of the new high reliability system which enables it to cope with it suitable for failure generating.

[0016]

[Means for Solving the Problem] In order to attain this purpose, the high reliability system of this invention It has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network. present -- business -- When a failure occurs on one of disks and the configuration of carrying out business by using normal another disk is taken When failure generating of an original disk is detected by a detection means to detect whether it is that the failure occurred on the original disk, and the detection means it corresponds to the node change by the change means which changes a standby system node to an employment condition, and the change means -- making -- till then -- present -- business -- it constitutes so that it may have the means for stopping which stops duplex system by stopping the node which was operating as a system node.

[0017] thus -- the high reliability system of this invention constituted -- present -- business -- if a failure occurs in the hard disk of the node which is operating as a system, while changing into an employment condition the node which is operating as a standby system -- till then -- present -- business -- by stopping the node which was operating as a system, it processes so that the actuation as duplex system may be stopped.

[0018] Since the node which newly performs operating processing by this can carry out business using the hard disk of a self-node, it is lost that the entry-of-data engine performance deteriorates of it, and it can prevent system-wide degradation. And inconvenient actuation is not performed after that from suspending the actuation as duplex system.

[0019] In order to attain this purpose, moreover, the high reliability system of this invention It has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network. present -- business -- When a failure occurs on one of disks and the configuration of carrying out business by using normal another disk is taken It constitutes so that it may have a detection means to detect whether it is that the failure occurred on the shadow disk, and the means for stopping which stops duplex system by stopping a standby node when failure generating of a shadow disk is detected by the detection means.

[0020] Thus, with the high reliability system of this invention constituted, if a failure occurs in the hard disk of the node which is operating as a standby system, it will process by stopping the node which is operating as a standby system, so that the actuation as duplex system may be stopped.

[0021] Thereby, the node which performs operating processing is being able to tell a user now about the standby system node not being prepared, and while it is lost that a user has anxiety to the dependability

of a system, it does not happen un-arranging [ that business will be succeeded as old operating data are also ], either.

[0022] In order to attain this purpose, moreover, the high reliability system of this invention It has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network. present -- business -- When a failure occurs on one of disks and the configuration of carrying out business by using normal another disk is taken present -- business, when failure generating of the network for shadows is detected by a detection means to detect whether it is that the failure occurred to the network for shadows prepared in order that a system node may access a shadow disk, and the detection means It constitutes so that it may have the means for stopping which stops duplex system by stopping a standby node.

[0023] Thus, with the high reliability system of this invention constituted, if a failure occurs to the network for shadows, it will process by stopping the node which is operating as a standby system, so that the actuation as duplex system may be stopped.

[0024] Thereby, the node which performs operating processing is being able to tell a user now about the standby system node not being prepared, and while it is lost that a user has anxiety to the dependability of a system, it does not happen un-arranging [ that business will be succeeded as old operating data are also ], either.

[0025] thus -- according to this invention -- present -- business -- it is normal, when it has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network and a failure occurs on one of disks -- already -- it is, when taking the configuration of carrying out business by using a disk, and while can cope with it now suitable for failure generating.

[0026]

[Embodiment of the Invention] Hereafter, this invention is explained to a detail according to the gestalt of operation.

[0027] The system configuration of the high reliability system with which this invention is applied to drawing 1 is illustrated.

[0028] As shown in this drawing, the high reliability system with which this invention is applied operating processing is performed -- present -- business -- the system node 1 -- present -- business -- with the standby system node 2 prepared as a standby system of the system node 1 present -- business -- the exclusive network 3 which connects between the system node 1 and the standby system nodes 2, and the business-use terminal 4 which a user operates -- present -- business, while connecting between the system node 1 and the standby system nodes 2 present -- business -- it consists of business-use networks 5 which connect between the system node 1/the standby system node 2, and the business-use terminals 4.

[0029] here -- the exclusive network 3 -- present -- business -- while being used as communication link Rhine in case the system node 1 accesses the hard disk of the standby system node 2 -- present -- business -- it is used as communication link Rhine of the notice of survival exchanged between the system node 1 and the standby system node 2, and has doubled preferably.

[0030] this -- present -- business -- the system node 1 is equipped with the hard disk 10 which functions as an original disk of mirroring, the application program 11 which performs operating processing using a hard disk 10, the network disk driver 12 for accessing the hard disk 20 of the standby system node 2, the mirror driver 13 which performs mirroring to hard disks 10 and 20, and the cluster control section 14 which performs the whole control processing.

[0031] the hard disk 20 on which this standby system node 2 functions as a shadow disk of mirroring on the other hand -- present -- business -- the application program 21 prepared for the application program 11 of the system node 1 by matching -- present -- business -- it has the network disk driver 22 for accessing the hard disk 10 of the system node 1, the mirror driver 23 which performs mirroring to hard disks 20 and 10, and the cluster control section 24 which performs the whole control processing.

[0032] drawing 2 -- present -- business -- the example of 1 operation gestalt of this invention which functions when a failure occurs in the hard disk 10 of the system node 1 is illustrated.

[0033] here -- the inside of drawing -- present -- business -- the system node 1 -- "the # node 1" and the standby system node 2 -- "the # node 2" -- present -- business -- "HD2" shows the hard disk 20 of "HD1" and the standby system node 2 for the hard disk 10 of the system node 1.

[0034] next, this drawing -- following -- present -- business -- when a failure occurs in the hard disk 10 of the system node 1, the processing which this invention performs is explained.

[0035] (1) present -- business -- suppose that a certain failure occurred in the hard disk 10 of the system node 1.

[0036] (2) the time of this failure having occurred -- present -- business, if the write-in demand of operating data is published that the application program 11 of the system node 1 should carry out operating processing (3) -- present -- business -- the mirror driver 13 of the system node 1 While writing in the operating data which have a write-in demand in the hard disk 10 of a self-node that mirroring should be performed, the operating data which have a write-in demand in the hard disk 20 of the standby system node 2 using the network disk driver 12 will be written in. this time -- present -- business -- the mirror driver 13 of the system node 1 is being unable to perform this writing to the hard disk 10 of a self-node, and detects that the failure has occurred in the hard disk 10 of a self-node.

[0037] (4) present -- business -- the mirror driver 13 of the system node 1 will notify that to the cluster control section 14 of a self-node, if it detects that the failure has occurred in the hard disk 10 of a self-node.

[0038] (5) present -- business -- if this notice is received, since the application program 11 of (6) self-node is stopped, the cluster control section 14 of the system node 1 will change a self-node from an employment condition to a idle state -- making -- (7) -- after it, it is stopping the mirror driver 13 of a self-node, and stop the actuation as duplex system.

[0039] (8) then -- present -- business -- the actuation as a system was suspended -- present -- business -- the cluster control section 14 of the system node 1 notifies a change in the employment condition to the cluster control section 24 of the standby system node 2 using the business-use network 5.

[0040] (9) if this notice is received, the cluster control section 24 of the standby system node 2 will change a self-node in the employment condition from a idle state -- making -- (10) -- while starting the mirror driver 23 of a self-node after it, resume business by starting the application program 21 of (11) self-node.

[0041] Here, when starting the mirror driver 23, only the hard disk 20 of a self-node is made accessible, and processing which changes the configuration information of mirroring is performed about the hard disk 10 which the failure generated to make it not access.

[0042] thus -- this invention -- present -- business -- if a failure occurs in the hard disk of the node which is operating as a system, while changing into an employment condition the node which is operating as a standby system -- till then -- present -- business -- by stopping the node which was operating as a system, it processes so that the actuation as duplex system may be stopped.

[0043] Since the node which newly performs operating processing by this can carry out business using the hard disk of a self-node, it is lost that the entry-of-data engine performance deteriorates of it, and it can prevent system-wide degradation. And inconvenient actuation is not performed after that from suspending the actuation as duplex system.

[0044] The example of 1 operation gestalt of this invention which functions on drawing 3 when a failure occurs in the hard disk 20 of the standby system node 2 is illustrated.

[0045] here -- the inside of drawing -- present -- business -- the system node 1 -- "the # node 1" and the standby system node 2 -- "the # node 2" -- present -- business -- "HD2" shows the hard disk 20 of "HD1" and the standby system node 2 for the hard disk 10 of the system node 1.

[0046] Next, when a failure occurs in the hard disk 20 of the standby system node 2 according to this drawing, the processing which this invention performs is explained.

[0047] (1) Suppose that a certain failure occurred in the hard disk 20 of the standby system node 2.

[0048] (2) the time of this failure having occurred -- present -- business, if the write-in demand of operating data is published that the application program 11 of the system node 1 should carry out operating processing (3) -- present -- business -- the mirror driver 13 of the system node 1 While writing

in the operating data which have a write-in demand in the hard disk 10 of a self-node that mirroring should be performed, the operating data which have a write-in demand in the hard disk 20 of the standby system node 2 using the network disk driver 12 will be written in. this time -- present -- business -- the mirror driver 13 of the system node 1 is being unable to perform this writing to the hard disk 20 of the standby system node 2, and detects that the failure has occurred in the hard disk 20 of the standby system node 2.

[0049] (4) present -- business -- the mirror driver 13 of the system node 1 will notify that to the cluster control section 14 of a self-node, if it detects that the failure occurred in the hard disk 20 of the standby system node 2.

[0050] (5) if this notice is received -- present -- business -- the cluster control section 14 of the system node 1 -- the business-use network 5 -- using -- the cluster control section 24 of the standby system node 2 -- receiving -- cut-off processing -- notifying -- (6) -- in response, the cluster control section 24 of the standby system node 2 is making it change from a standby condition to a idle state, and stops the actuation as duplex system.

[0051] Thus, by this invention, if a failure occurs in the hard disk of the node which is operating as a standby system, it will process by stopping the node which is operating as a standby system, so that the actuation as duplex system may be stopped.

[0052] Thereby, the node which performs operating processing is being able to tell a user now about the standby system node not being prepared, and while it is lost that a user has anxiety to the dependability of a system, it does not happen un-arranging [ that business will be succeeded as old operating data are also ], either.

[0053] The example of 1 operation gestalt of this invention which functions on drawing 4 when a failure occurs to the exclusive network 3 is illustrated.

[0054] here -- the inside of drawing -- present -- business -- the system node 1 -- "the # node 1" and the standby system node 2 -- "the # node 2" -- present -- business -- "HD2" shows the hard disk 20 of "HD1" and the standby system node 2 for the hard disk 10 of the system node 1.

[0055] Next, when a failure occurs to the exclusive network 3 according to this drawing, the processing which this invention performs is explained.

[0056] (1) Suppose that a certain failure occurred to the exclusive network 3.

[0057] (2) the time of this failure having occurred -- present -- business, if the write-in demand of operating data is published that the application program 11 of the system node 1 should carry out operating processing (3) -- present -- business -- the mirror driver 13 of the system node 1 While writing in the operating data which have a write-in demand in the hard disk 10 of a self-node that mirroring should be performed, the operating data which have a write-in demand in the hard disk 20 of the standby system node 2 using the network disk driver 12 will be written in. this time -- present -- business -- the mirror driver 13 of the system node 1 is being unable to perform this writing to the hard disk 20 of the standby system node 2, and detects that the failure has occurred to the exclusive network 3.

[0058] (4) present -- business -- the mirror driver 13 of the system node 1 will notify that to the cluster control section 14 of a self-node, if it detects that the failure occurred to the exclusive network 3.

[0059] (5) if this notice is received -- present -- business -- the cluster control section 14 of the system node 1 This is received. the business-use network 5 -- using -- the cluster control section 24 of the standby system node 2 -- receiving -- cut-off processing -- notifying -- (6) -- the cluster control section 24 of the standby system node 2 While making it change from a standby condition to a idle state, the actuation as duplex system is stopped by directing to separate the hard disk 20 of a self-node from a mirror set to the mirror driver 23 of (7) self-node.

[0060] (8) And if these directions are received, the mirror driver 23 will be updating the configuration information of the mirror set memorized by the hard disk 20 of a self-node in the form which this hard disk's 20 cannot use, and will separate the hard disk 20 of a self-node from a mirror set.

[0061] Thus, by this invention, if a failure occurs to the exclusive network 3, it will process by stopping the node which is operating as a standby system, so that the actuation as duplex system may be suspended.

[0062] Thereby, the node which performs operating processing is being able to tell a user now about the standby system node not being prepared, and while it is lost that a user has anxiety to the dependability of a system, it does not happen un-arranging [ that business will be succeeded as old operating data are also ], either.

[0063]

[Effect of the Invention] it explained above -- as -- the high reliability system of this invention -- present -- business -- if a failure occurs in the hard disk of the node which is operating as a system, while changing into an employment condition the node which is operating as a standby system -- till then -- present -- business -- by stopping the node which was operating as a system, it processes so that the actuation as duplex system may be stopped.

[0064] Since the node which newly performs operating processing by this can carry out business using the hard disk which a self-node has, it is lost that the entry-of-data engine performance deteriorates of it, and it can prevent system-wide degradation. And inconvenient actuation is not performed after that from suspending the actuation as duplex system.

[0065] Moreover, with the high reliability system of this invention, if a failure occurs in the hard disk of the node which is operating as a standby system, it will process by stopping the node which is operating as a standby system, so that the actuation as duplex system may be stopped.

[0066] Thereby, the node which performs operating processing is being able to tell a user now about the standby system node not being prepared, and while it is lost that a user has anxiety to the dependability of a system, it does not happen un-arranging [ that business will be succeeded as old operating data are also ], either.

[0067] Moreover, with the high reliability system of this invention, if a failure occurs to the network for shadows prepared for access of a shadow disk, it will process by stopping the node which is operating as a standby system, so that the actuation as duplex system may be stopped.

[0068] Thereby, the node which performs operating processing is being able to tell a user now about the standby system node not being prepared, and while it is lost that a user has anxiety to the dependability of a system, it does not happen un-arranging [ that business will be succeeded as old operating data are also ], either.

[0069] thus -- according to this invention -- present -- business -- it is normal, when it has the mirror set function which uses the local disk of a system node as an original disk, and uses as a shadow disk the local disk of a standby system node connected through a network and a failure occurs on one of disks -- already -- it is, when taking the configuration of carrying out business by using a disk, and while can cope with it now suitable for failure generating.

---

[Translation done.]

**\* NOTICES \***

**JPO and NCIPi are not responsible for any damages caused by the use of this translation.**

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

**DESCRIPTION OF DRAWINGS**

---

[Brief Description of the Drawings]

[Drawing 1] It is the system configuration Fig. of the high reliability system with which this invention is applied.

[Drawing 2] It is the example of 1 operation gestalt of this invention.

[Drawing 3] It is the example of 1 operation gestalt of this invention.

[Drawing 4] It is the example of 1 operation gestalt of this invention.

[Drawing 5] It is the explanatory view of the high reliability system with which this invention is applied.

[Description of Notations]

1 Present -- Business -- System Node

2 Standby System Node

3 Exclusive Network

4 Business-use Terminal

5 Business-use Network

10 20 Hard disk

11 21 Application program

12 22 Network disk driver

13 23 Mirror driver

14 24 Cluster control section

---

[Translation done.]

(19)日本特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-123406

(P2002-123406A)

(43)公開日 平成14年4月26日(2002.4.26)

(51)Int.Cl.	識別記号	F I	テマコード(参考)
G 0 6 F 11/20	3 1 0	G 0 6 F 11/20	3 1 0 E 5 B 0 1 8
3/06	3 0 4	3/06	3 0 4 B 5 B 0 3 4
12/16	3 1 0	12/16	3 1 0 M 5 B 0 6 5

審査請求 未請求 請求項の数3 OL (全 9 頁)

(21)出願番号 特願2000-316213(P2000-316213)

(22)出願日 平成12年10月17日(2000.10.17)

(71)出願人 000136136

株式会社ビーエフユー

石川県河北郡宇ノ気町宇野気ヌ98番地の  
2

(72)発明者 河合 桂介

石川県河北郡宇ノ気町宇野気ヌ98番地の  
2 株式会社ビーエフユー内

(72)発明者 島田 一祥

石川県河北郡宇ノ気町宇野気ヌ98番地の  
2 株式会社ビーエフユー内

(74)代理人 100095072

弁理士 岡田 光由 (外1名)

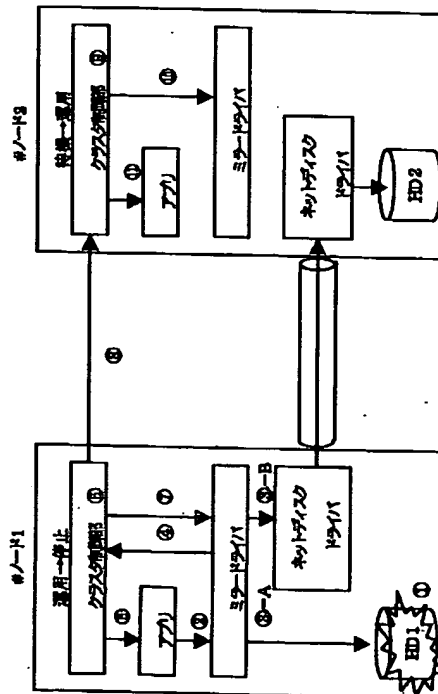
最終頁に続く

(54)【発明の名称】 高信頼性システム

(57)【要約】

【課題】本発明は、現用系ノードのディスクをオリジナルとし、ネットワークを介して接続される待機系ノードのディスクをシャドウとするミラーリング構成を採用するときに、障害発生に適切に対処できるようにすることを目的とする。

【解決手段】現用系ノードのディスクに障害が発生すると、待機系として動作しているノードを運用状態にするとともに、それまで現用系として動作していたノードを停止させることで、二重化システムとしての動作を停止させるように処理する。また、待機系ノードのディスクに障害が発生すると、待機系として動作しているノードを停止させることで、二重化システムとしての動作を停止させるように処理する。また、シャドウディスクへのアクセス用に用意されるネットワークに障害が発生すると、待機系として動作しているノードを停止させることで、二重化システムとしての動作を停止させるように処理する。



## 【特許請求の範囲】

【請求項1】 現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行する高信頼性システムにおいて、

上記オリジナルディスクに障害が発生したのか否かを検出する手段と、

上記オリジナルディスクの障害発生が検出されるときに、待機系ノードを運用状態に切り替える手段と、

上記ノード切り替えに対応させて、それまで現用系ノードとして動作していたノードを停止させることで二重化システムを停止させる手段とを備えることを、

特徴とする高信頼性システム。

【請求項2】 現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行する高信頼性システムにおいて、

上記シャドウディスクに障害が発生したのか否かを検出する手段と、

上記シャドウディスクの障害発生が検出されるときに、待機ノードを停止させることで二重化システムを停止させる手段とを備えることを、

特徴とする高信頼性システム。

【請求項3】 現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行する高信頼性システムにおいて、

現用系ノードが上記シャドウディスクにアクセスするために用意されるシャドウ用ネットワークに障害が発生したのか否かを検出する手段と、

上記シャドウ用ネットワークの障害発生が検出されるときに、待機ノードを停止させることで二重化システムを停止させる手段とを備えることを、

特徴とする高信頼性システム。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常

なもう一方のディスクを使用することで業務を遂行する高信頼性システムに関し、特に、障害発生に適切に対処できるようにする高信頼性システムに関する。

## 【0002】

【従来の技術】業務を行うサーバと、そのサーバの障害に備えて待機するサーバとから構成される高信頼性システムが広く用いられている。前者は現用系ノードと呼ばれ、後者は待機系ノードと呼ばれている。

【0003】現用系ノードが何らかの障害によりダウンした場合、待機系ノードが業務の引き継ぎを行う。この業務の引き継ぎを行うために、業務データを現用系ノードから待機系ノードへ引き継ぐ必要がある。

【0004】これまで、この業務データの引き継ぎのために、現用系ノードと待機系ノードとの間に、SCSIインタフェースなどで接続される共用ディスク装置を用意するという構成を採っていた。

【0005】しかしながら、共用ディスク装置は極めて高価なものである。そこで、現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能(RAID1の機能)を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行するという構成を採る高信頼性システムが利用されつつある。

【0006】この高信頼性システムでは、図5に示すように、現用系ノードは、オリジナルディスクに業務データを書き込むときに、ネットディスクドライバを使ってシャドウディスクにその業務データを書き込むことで、ミラーリングを実行するという構成を採っている。

【0007】そして、現用系ノードは、オリジナルディスクが正常である場合には、オリジナルディスクにアクセスすることで業務を遂行し、オリジナルディスクに障害が発生する場合には、ネットディスクドライバを使ってシャドウディスクにアクセスすることで業務を遂行するという構成を採っている。

## 【0008】

【発明が解決しようとする課題】しかしながら、この従来技術には、次のような問題点がある。

【0009】すなわち、オリジナルディスクに障害が発生するときに、現用系ノードは、業務データの入出力をすべてシャドウディスクに対して行うことから、データの入力性能が低下し、これがために、システム全体の性能低下が発生するという問題点がある。

【0010】また、シャドウディスクに障害が発生するときに、この従来技術では、待機系ノードを停止させるという構成を採っていない。

【0011】これから、シャドウディスクに障害が発生した後、現用系ノードがオリジナルディスクを使って業務を遂行していくときに、現用系ノードがシステムダウ

ンすることが起こると、ユーザは、待機系ノードに業務が引き継がれていく筈であるのにそれが実行されないことで、システムの信頼性に対して不安を持つという問題点がある。

【0012】しかも、この現用系ノードのシステムダウンの際に、障害箇所を避けることでシャドウディスクを使用することが可能になることがあっても、そのシャドウディスクの業務データは更新されていないことから、新現用系ノード（そのシャドウディスクを持つノード）は、最新の業務データを引き継ぐことができないという問題点もある。

【0013】また、現用系ノードがシャドウディスクにアクセスするために用意されるネットワーク（図5の専用ネットワーク）に障害が発生するときにも、この従来技術では、待機系ノードを停止させるという構成を採っていない。

【0014】これから、シャドウディスクに障害が発生するときの問題点と同様の問題点がある。

【0015】本発明はかかる事情に鑑みてなされたものであって、現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行するという構成を採るときにあって、障害発生に適切に対処できるようにする新たな高信頼性システムの提供を目的とする。

【0016】

【課題を解決するための手段】この目的を達成するために、本発明の高信頼性システムは、現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行するという構成を採るときに、オリジナルディスクに障害が発生したのか否かを検出する検出手段と、検出手段によりオリジナルディスクの障害発生が検出されるときに、待機系ノードを運用状態に切り替える切替手段と、切替手段によるノード切り替えに対応させて、それまで現用系ノードとして動作していたノードを停止させることで二重化システムを停止させる停止手段とを備えるように構成する。

【0017】このように構成される本発明の高信頼性システムでは、現用系として動作しているノードのハードディスクに障害が発生すると、待機系として動作しているノードを運用状態にするとともに、それまで現用系として動作していたノードを停止させることで、二重化システムとしての動作を停止させるように処理する。

【0018】これにより、新たに業務処理を実行するノ

ードは、自ノードのハードディスクを使って業務を遂行できるようになるので、データの入力性能が劣化することがなくなり、システム全体の性能低下を防止できるようになる。そして、二重化システムとしての動作を停止することから、その後、不都合な動作を行うことがない。

【0019】また、この目的を達成するために、本発明の高信頼性システムは、現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行するという構成を採るときに、シャドウディスクに障害が発生したのか否かを検出する検出手段と、検出手段によりシャドウディスクの障害発生が検出されるときに、待機ノードを停止させることで二重化システムを停止させる停止手段とを備えるように構成する。

【0020】このように構成される本発明の高信頼性システムでは、待機系として動作しているノードのハードディスクに障害が発生すると、待機系として動作しているノードを停止させることで、二重化システムとしての動作を停止させるように処理する。

【0021】これにより、業務処理を実行するノードは、待機系ノードが用意されていないことをユーザに知らせることができるようになることで、ユーザがシステムの信頼性に対して不安を持つということがなくなるとともに、古い業務データでもって業務が引き継がれてしまうという不都合も起こらない。

【0022】また、この目的を達成するために、本発明の高信頼性システムは、現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行するという構成を採るときに、現用系ノードがシャドウディスクにアクセスするために用意されるシャドウ用ネットワークに障害が発生したのか否かを検出する検出手段と、検出手段によりシャドウ用ネットワークの障害発生が検出されるときに、待機ノードを停止させることで二重化システムを停止させる停止手段とを備えるように構成する。

【0023】このように構成される本発明の高信頼性システムでは、シャドウ用ネットワークに障害が発生すると、待機系として動作しているノードを停止させることで、二重化システムとしての動作を停止させるように処理する。

【0024】これにより、業務処理を実行するノードは、待機系ノードが用意されていないことをユーザに知らせることができるようになることで、ユーザがシステ

ムの信頼性に対して不安を持つということがなくなるとともに、古い業務データでもって業務が引き継がれてしまうという不都合も起こらない。

【0025】このようにして、本発明によれば、現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行するという構成を採るときにあって、障害発生に適切に対処できるようになる。

【0026】

【発明の実施の形態】以下、実施の形態に従って本発明を詳細に説明する。

【0027】図1に、本発明の適用される高信頼性システムのシステム構成を図示する。

【0028】この図に示すように、本発明の適用される高信頼性システムは、業務処理を実行する現用系ノード1と、現用系ノード1の待機系として用意される待機系ノード2と、現用系ノード1と待機系ノード2との間を接続する専用ネットワーク3と、ユーザの操作する業務用端末4と、現用系ノード1と待機系ノード2との間を接続するとともに、現用系ノード1/待機系ノード2と業務用端末4との間を接続する業務用ネットワーク5とで構成されている。

【0029】ここで、専用ネットワーク3は、現用系ノード1が待機系ノード2のハードディスクにアクセスするときの通信ラインとして用いられるとともに、現用系ノード1と待機系ノード2との間でやり取りされる生存通知の通信ラインとして用いられるものであり、好ましくは二重化されている。

【0030】この現用系ノード1は、ミラーリングのオリジナルディスクとして機能するハードディスク10と、ハードディスク10を使って業務処理を実行するアプリケーションプログラム11と、待機系ノード2のハードディスク20にアクセスするためのネットディスクドライバ12と、ハードディスク10、20に対するミラーリングを実行するミラードライバ13と、全体の制御処理を実行するクラスタ制御部14とを備える。

【0031】一方、この待機系ノード2は、ミラーリングのシャドウディスクとして機能するハードディスク20と、現用系ノード1のアプリケーションプログラム11に対応付けて用意されるアプリケーションプログラム21と、現用系ノード1のハードディスク10にアクセスするためのネットディスクドライバ22と、ハードディスク20、10に対するミラーリングを実行するミラードライバ23と、全体の制御処理を実行するクラスタ制御部24とを備える。

【0032】図2に、現用系ノード1のハードディスク10に障害が発生するときに機能する本発明の一実施形

態例を図示する。

【0033】ここで、図中、現用系ノード1を「#ノード1」、待機系ノード2を「#ノード2」、現用系ノード1のハードディスク10を「HD1」、待機系ノード2のハードディスク20を「HD2」で示している。

【0034】次に、この図に従って、現用系ノード1のハードディスク10に障害が発生したときに、本発明が実行する処理について説明する。

【0035】(1) 現用系ノード1のハードディスク10に何らかの障害が発生したとする。

【0036】(2) この障害が発生しているときに、現用系ノード1のアプリケーションプログラム11が業務処理を遂行すべく業務データの書込要求を発行すると、

(3) 現用系ノード1のミラードライバ13は、ミラーリングを実行すべく、自ノードのハードディスク10に書込要求のある業務データの書き込みを行うとともに、ネットディスクドライバ12を使って待機系ノード2のハードディスク20に書込要求のある業務データの書き込みを行うことになる。このとき、現用系ノード1のミラードライバ13は、自ノードのハードディスク10にはこの書き込みを実行できないことで、自ノードのハードディスク10に障害が発生していることを検出する。

【0037】(4) 現用系ノード1のミラードライバ13は、自ノードのハードディスク10に障害が発生していることを検出すると、その旨を自ノードのクラスタ制御部14に通知する。

【0038】(5) 現用系ノード1のクラスタ制御部14は、この通知を受け取ると、(6) 自ノードのアプリケーションプログラム11を停止させてから、自ノードを運用状態から停止状態に移転させ、(7) それに続けて、自ノードのミラードライバ13を停止させることで、二重化システムとしての動作を停止させる。

【0039】(8) 続いて、現用系としての動作を停止した現用系ノード1のクラスタ制御部14は、業務用ネットワーク5を使って、待機系ノード2のクラスタ制御部24に対して運用状態への切り替えを通知する。

【0040】(9) この通知を受け取ると、待機系ノード2のクラスタ制御部24は、自ノードを停止状態から運用状態に移転させ、(10) それに続けて、自ノードのミラードライバ23を起動するとともに、(11) 自ノードのアプリケーションプログラム21を起動することで業務を再開する。

【0041】ここで、ミラードライバ23を起動するときには、自ノードのハードディスク20のみをアクセス可能とし、障害の発生したハードディスク10についてはアクセスしないようにすべく、ミラーリングの構成情報を変更する処理を行う。

【0042】このようにして、本発明では、現用系として動作しているノードのハードディスクに障害が発生すると、待機系として動作しているノードを運用状態にす

るとともに、それまで現用系として動作していたノードを停止させることで、二重化システムとしての動作を停止させるように処理する。

【0043】これにより、新たに業務処理を実行するノードは、自ノードのハードディスクを使って業務を遂行できるようになるので、データの入力性能が劣化するということがなくなり、システム全体の性能低下を防止できるようになる。そして、二重化システムとしての動作を停止することから、その後、不都合な動作を行うことがない。

【0044】図3に、待機系ノード2のハードディスク20に障害が発生するときに機能する本発明の一実施形態例を図示する。

【0045】ここで、図中、現用系ノード1を「#ノード1」、待機系ノード2を「#ノード2」、現用系ノード1のハードディスク10を「HD1」、待機系ノード2のハードディスク20を「HD2」で示している。

【0046】次に、この図に従って、待機系ノード2のハードディスク20に障害が発生するときに、本発明が実行する処理について説明する。

【0047】(1) 待機系ノード2のハードディスク20に何らかの障害が発生したとする。

【0048】(2) この障害が発生しているときに、現用系ノード1のアプリケーションプログラム11が業務処理を遂行すべく業務データの書込要求を発行すると、

(3) 現用系ノード1のミラードライバ13は、ミラーリングを実行すべく、自ノードのハードディスク10に書込要求のある業務データの書き込みを行うとともに、ネットディスクドライバ12を使って待機系ノード2のハードディスク20に書込要求のある業務データの書き込みを行うことになる。このとき、現用系ノード1のミラードライバ13は、待機系ノード2のハードディスク20にはこの書き込みを実行できないことで、待機系ノード2のハードディスク20に障害が発生していることを検出する。

【0049】(4) 現用系ノード1のミラードライバ13は、待機系ノード2のハードディスク20に障害が発生したことを検出すると、その旨を自ノードのクラスタ制御部14に通知する。

【0050】(5) この通知を受け取ると、現用系ノード1のクラスタ制御部14は、業務用ネットワーク5を使って、待機系ノード2のクラスタ制御部24に対して切り捨て処理を通知し、(6) これを受けて、待機系ノード2のクラスタ制御部24は、待機状態から停止状態へ遷移させることで、二重化システムとしての動作を停止させる。

【0051】このようにして、本発明では、待機系として動作しているノードのハードディスクに障害が発生すると、待機系として動作しているノードを停止させることで、二重化システムとしての動作を停止させるように

処理する。

【0052】これにより、業務処理を実行するノードは、待機系ノードが用意されていないことをユーザに知らせることができるようになることで、ユーザがシステムの信頼性に対して不安を持つということがなくなるとともに、古い業務データでもって業務が引き継がれてしまうという不都合も起こらない。

【0053】図4に、専用ネットワーク3に障害が発生するときに機能する本発明の一実施形態例を図示する。

【0054】ここで、図中、現用系ノード1を「#ノード1」、待機系ノード2を「#ノード2」、現用系ノード1のハードディスク10を「HD1」、待機系ノード2のハードディスク20を「HD2」で示している。

【0055】次に、この図に従って、専用ネットワーク3に障害が発生するときに、本発明が実行する処理について説明する。

【0056】(1) 専用ネットワーク3に何らかの障害が発生したとする。

【0057】(2) この障害が発生しているときに、現用系ノード1のアプリケーションプログラム11が業務処理を遂行すべく業務データの書込要求を発行すると、

(3) 現用系ノード1のミラードライバ13は、ミラーリングを実行すべく、自ノードのハードディスク10に書込要求のある業務データの書き込みを行うとともに、ネットディスクドライバ12を使って待機系ノード2のハードディスク20に書込要求のある業務データの書き込みを行うことになる。このとき、現用系ノード1のミラードライバ13は、待機系ノード2のハードディスク20にはこの書き込みを実行できないことで、専用ネットワーク3に障害が発生していることを検出する。

【0058】(4) 現用系ノード1のミラードライバ13は、専用ネットワーク3に障害が発生したことを検出すると、その旨を自ノードのクラスタ制御部14に通知する。

【0059】(5) この通知を受け取ると、現用系ノード1のクラスタ制御部14は、業務用ネットワーク5を使って、待機系ノード2のクラスタ制御部24に対して切り捨て処理を通知し、(6) これを受けて、待機系ノード2のクラスタ制御部24は、待機状態から停止状態へ遷移させるとともに、(7) 自ノードのミラードライバ23に対して、自ノードのハードディスク20をミラーセットから切り離すことを指示することで、二重化システムとしての動作を停止させる。

【0060】(8) そして、この指示を受け取ると、ミラードライバ23は、自ノードのハードディスク20に記憶されるミラーセットの構成情報をこのハードディスク20が使用できない形に更新することで、自ノードのハードディスク20をミラーセットから切り離す。

【0061】このようにして、本発明では、専用ネットワーク3に障害が発生すると、待機系として動作してい

るノードを停止させることで、二重化システムとしての動作を停止するように処理する。

【0062】これにより、業務処理を実行するノードは、待機系ノードが用意されていないことをユーザに知らせることができるようになることで、ユーザがシステムの信頼性に対して不安を持つということがなくなるとともに、古い業務データでもって業務が引き継がれてしまうという不都合も起こらない。

【0063】

【発明の効果】以上説明したように、本発明の高信頼性システムでは、現用系として動作しているノードのハードディスクに障害が発生すると、待機系として動作しているノードを運用状態にするとともに、それまで現用系として動作していたノードを停止させることで、二重化システムとしての動作を停止させるように処理する。

【0064】これにより、新たに業務処理を実行するノードは、自ノードの持つハードディスクを使って業務を遂行できるようになるので、データの入力性能が劣化するということがなくなり、システム全体の性能低下を防止できるようになる。そして、二重化システムとしての動作を停止することから、その後、不都合な動作を行うことがない。

【0065】また、本発明の高信頼性システムでは、待機系として動作しているノードのハードディスクに障害が発生すると、待機系として動作しているノードを停止させることで、二重化システムとしての動作を停止させるように処理する。

【0066】これにより、業務処理を実行するノードは、待機系ノードが用意されていないことをユーザに知らせることができるようになることで、ユーザがシステムの信頼性に対して不安を持つということがなくなるとともに、古い業務データでもって業務が引き継がれてしまうという不都合も起こらない。

【0067】また、本発明の高信頼性システムでは、シャドウディスクのアクセス用に用意されるシャドウ用ネットワークに障害が発生すると、待機系として動作して

いるノードを停止させることで、二重化システムとしての動作を停止させるように処理する。

【0068】これにより、業務処理を実行するノードは、待機系ノードが用意されていないことをユーザに知らせることができるようになることで、ユーザがシステムの信頼性に対して不安を持つということがなくなるとともに、古い業務データでもって業務が引き継がれてしまうという不都合も起こらない。

【0069】このようにして、本発明によれば、現用系ノードのローカルディスクをオリジナルディスクとし、ネットワークを介して接続される待機系ノードのローカルディスクをシャドウディスクとするミラーセット機能を有して、いずれか一方のディスクに障害が発生する場合に、正常なもう一方のディスクを使用することで業務を遂行するという構成を採るときにあって、障害発生に適切に対処できるようになる。

【図面の簡単な説明】

【図1】本発明の適用される高信頼性システムのシステム構成図である。

【図2】本発明の一実施形態例である。

【図3】本発明の一実施形態例である。

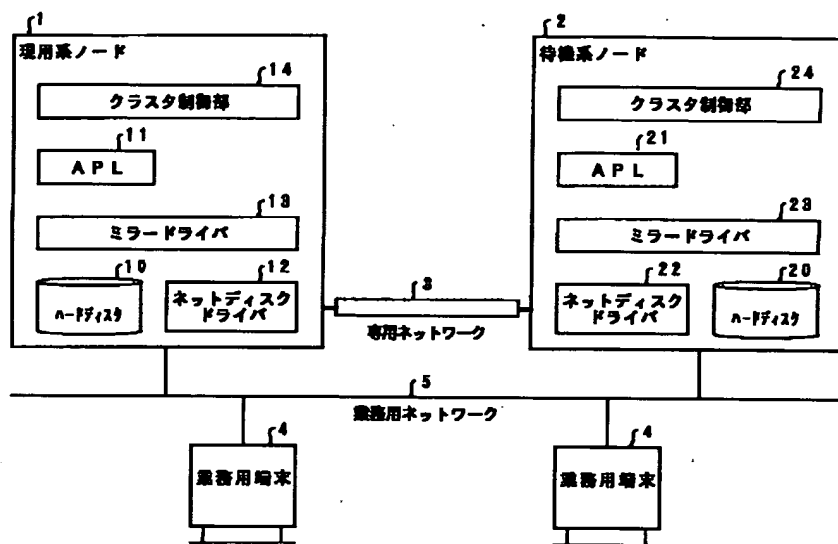
【図4】本発明の一実施形態例である。

【図5】本発明の適用される高信頼性システムの説明図である。

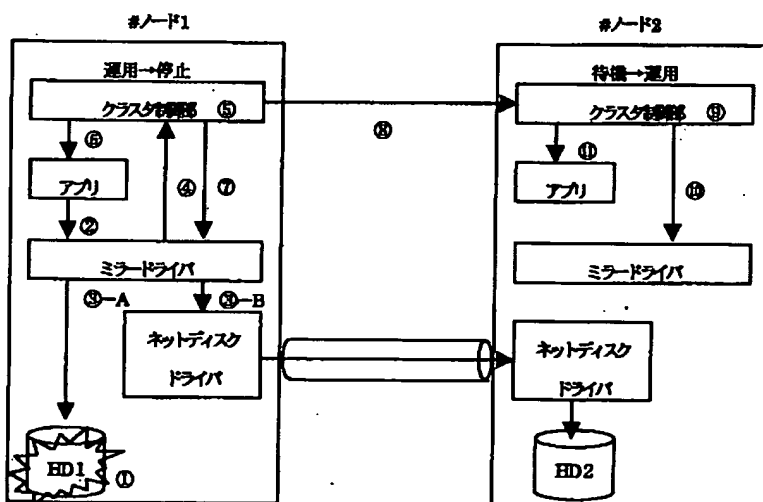
【符号の説明】

1	現用系ノード
2	待機系ノード
3	専用ネットワーク
4	業務用端末
5	業務用ネットワーク
10, 20	ハードディスク
11, 21	アプリケーションプログラム
12, 22	ネットディスクドライバ
13, 23	ミラードライバ
14, 24	クラスタ制御部

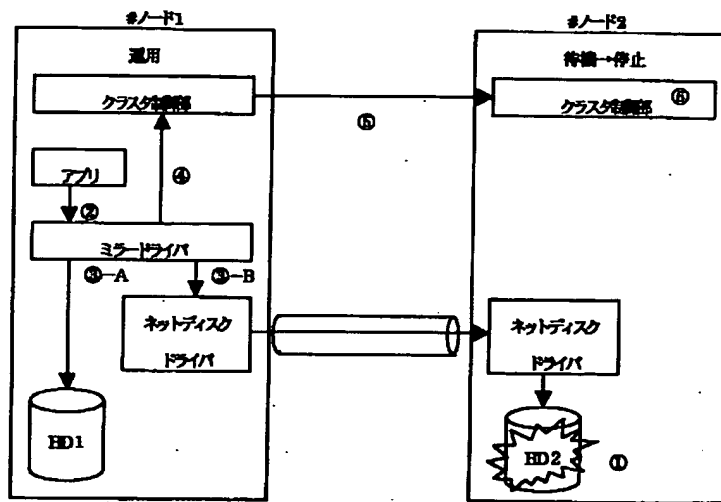
【図1】



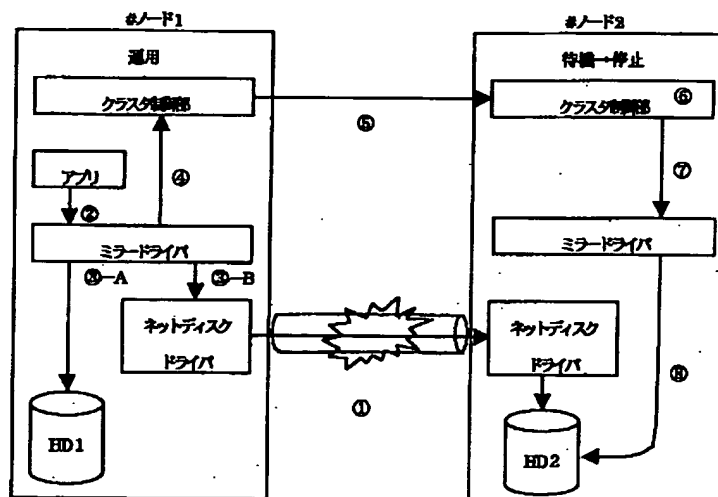
【図2】



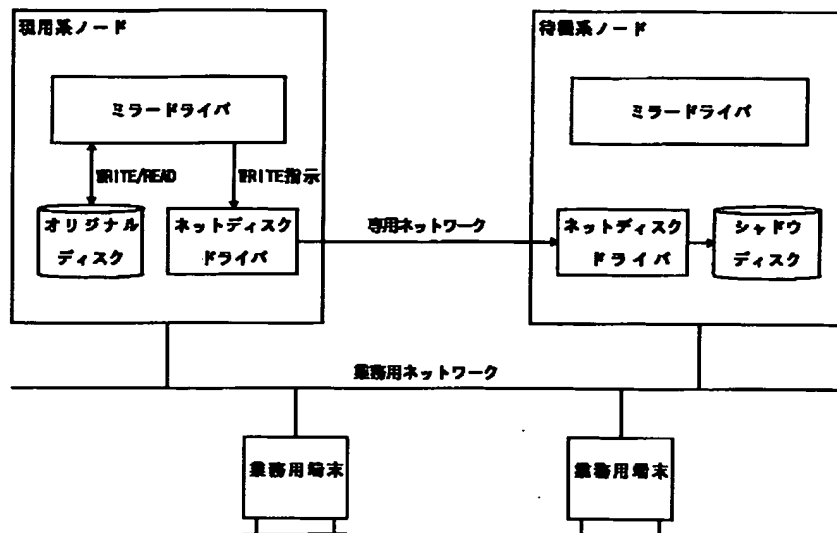
【図3】



【図4】



【図5】



フロントページの続き

(72)発明者 今井 秀和  
石川県河北郡宇ノ気町宇野気ヌ98番地の  
2 株式会社ビーエフユー内

Fターム(参考) 5B018 GA06 HA04 HA05 KA14 MA14  
QA20  
5B034 BB02 CC02  
5B065 BA01 EA12